

On McDiarmid's inequality for Hamming distance

Xiequan Fan*

Abstract

We improve the rate function of McDiarmid's inequality for Hamming distance. In particular, applying our result to the separately Lipschitz functions of independent random variables, we also refine the convergence rate function of McDiarmid's inequality around a median. Moreover, a non-uniform bound for the distance between the medians and the mean is also given. We also give some extensions of McDiarmid's inequalities to the case of nonnegative functionals of dependent random variables.

Keywords: McDiarmid's inequality; Concentration inequalities; Median.

AMS 2010 Subject Classification: 60E15.

Submitted to ECP on September ., 2016, final version accepted on December ., 2016.

1 Introduction

Let $\alpha = (\alpha_1, \dots, \alpha_n) \in [0, \infty)^n$ be an n -vector of non-negative real numbers. Recall that the L_2 norm is given by

$$\|\alpha\| = \left(\sum_{i=1}^n \alpha_i^2 \right)^{1/2},$$

and we call α a unit vector if $\|\alpha\| = 1$. For points $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ in \mathbf{E}^n , the α -Hamming distance $d_\alpha(x, y)$ is defined as

$$d_\alpha(x, y) = \sum_{x_i \neq y_i} \alpha_i.$$

In particular, when $\alpha_i = 1/\sqrt{n}$, the α -Hamming distance is just the unit Hamming distance. For a subset $A \subset \mathbf{E}^n$, we define

$$d_\alpha(x, A) = \inf_{y \in A} d_\alpha(x, y).$$

All over the paper, we assume that \mathbf{E} is a separable spaces such that $d_\alpha(x, A)$ is measurable.

McDiarmid proved the following concentration inequality for α -Hamming distance (see Theorem 3.6 of [4]). Let X be an \mathbf{E}^n -valued vector of independent random variables. Assume that $\|\alpha\| = 1$. Then, for any set $A \subset \mathbf{E}^n$ and any $t > 0$,

$$\mathbb{P}(d_\alpha(X, A) \geq t) \mathbb{P}(X \in A) \leq e^{-t^2/2}. \quad (1.1)$$

*Center for Applied Mathematics, Tianjin University, Tianjin 300072, China
 E-mail: fanxiequan@hotmail.com

Such result is very useful when one wants to evaluate the concentration around a median. To illustrate its application, consider the following example. All over the paper, let f be a function defined on \mathbf{E}^n . Assume that the function f satisfies

$$|f(x) - f(x')| \leq d_\alpha(x, x'). \quad (1.2)$$

Let m be a median of $f(X)$, that is $\mathbb{P}(f(X) \geq m) \geq \frac{1}{2}$ and $\mathbb{P}(f(X) \leq m) \geq \frac{1}{2}$. By taking $A = \{X \in \mathbf{E}^n : f(X) \leq m\}$, McDiarmid's inequality (1.1) implies the following concentration inequalities around a median: for any $t > 0$,

$$\mathbb{P}(f(X) - m \geq t) \leq 2e^{-\frac{1}{2}t^2} \quad (1.3)$$

and

$$\mathbb{P}(f(X) - m \leq -t) \leq 2e^{-\frac{1}{2}t^2}. \quad (1.4)$$

Denote by μ the mean value of the random variable $f(X)$, that is $\mu = \mathbb{E}[f(X)]$. By Lemma 4.6 of McDiarmid [4], the inequalities (1.3) and (1.4) together implies a uniform bound between the medians and the mean:

$$|m - \mu| \leq \sqrt{2\pi}. \quad (1.5)$$

McDiarmid (see Theorem 3.1 of [4]) also proved the following exponential moment inequality: for any $\lambda \geq 0$,

$$\mathbb{E}[e^{\lambda(f(X) - \mu)}] \leq \exp\left\{\frac{\lambda^2}{8}\right\}. \quad (1.6)$$

The last inequality together with the classical Bernstein inequality

$$\mathbb{P}(f(X) - \mu \geq t) \leq \inf_{\lambda \geq 0} e^{-\lambda t} \mathbb{E}[e^{\lambda(f(X) - \mu)}]$$

implies the following concentration inequality around the mean (instead of a median): for any $t > 0$,

$$\mathbb{P}(f(X) - \mu \geq t) \leq e^{-2t^2}. \quad (1.7)$$

See also Rio [6] for a recent improvement of (1.7), such that $\mathbb{P}(f(X) - \mu \geq t) = 0$ for any $t > \sum_{i=1}^n \alpha_i$. By (1.7), we get for any $t > |m - \mu|$,

$$\mathbb{P}(f(X) - m \geq t) \leq e^{-2(t - |m - \mu|)^2}. \quad (1.8)$$

The last inequality shows that for large t 's, the constant $\frac{1}{2}$ in the inequalities (1.3) and (1.4) can be improved. However, a deficiency of (1.8) is that comparing to (1.3) and (1.4), inequality (1.8) does not hold for $t \in (0, |m - \mu|]$. Moreover, in applications we need to give a bound for $|m - \mu|$. Obviously, this bound should be as small as possible.

The scope of the paper is to fill the deficiency of (1.8), and establish a bound sharper than (1.5). To achieve this scope, we give an improvement on (1.1): for any set $A \subset \mathbf{E}^n$ and any $t > 0$,

$$\mathbb{P}(d_\alpha(X, A) \geq t) \mathbb{P}(X \in A) \leq e^{-t^2}. \quad (1.9)$$

It is obvious that the r.h.s. of (1.9) is much smaller than the one of (1.1) due to the fact that the constant $\frac{1}{2}$ has been improved to 1. Using (1.9), we obtain an improvement on (1.3): for any $t > 0$,

$$\mathbb{P}(f(X) - m \geq t) \leq 2e^{-t^2}. \quad (1.10)$$

We also prove the following bound for $|m - \mu|$:

$$|\mu - m| \leq (2\rho + \sqrt{\pi/2})e^{-2\rho^2} < 1.5503, \quad (1.11)$$

where $\rho = \mathbb{E}[d_\alpha(X, A)]$ with $A = \{Y : f(Y) \leq m\}$. Since $1.5503 < \sqrt{2\pi}$, inequality (1.11) is tighter than (1.5). In particular, (1.11) implies a non-uniform bound between the medians and the mean, which shows that if $\rho \rightarrow \infty$, then the distance between the medians and the mean tends to 0 in an exponentially decaying rate.

This paper is organized as follows. Our main results are stated and discussed in Section 2. In Section 3, we extend the McDiarmid inequalities to the case of nonnegative functionals of *dependent* random variables. We prove our theorems in Section 4.

2 On McDiarmid's inequalities

In the following theorem, we give a stronger version of McDiarmid's inequality (1.1).

Theorem 2.1. *Let X be an \mathbb{E}^n -valued vector of independent random variables. Assume that $\|\alpha\| = 1$. Then for any set $A \subset \mathbb{E}^n$ and any $t > 0$,*

$$\mathbb{P}(d_\alpha(X, A) \geq t) \mathbb{P}(X \in A) \leq e^{-h(t)}, \quad (2.1)$$

where

$$h(t) = \begin{cases} 2(\mathbb{E}[d_\alpha(X, A)])^2, & \text{if } 0 < t < \mathbb{E}[d_\alpha(X, A)], \\ t^2 + (t - 2\mathbb{E}[d_\alpha(X, A)])^2, & \text{if } t \geq \mathbb{E}[d_\alpha(X, A)]. \end{cases} \quad (2.2)$$

In particular, it implies that for any $t > 0$,

$$\mathbb{P}(d_\alpha(X, A) \geq t) \mathbb{P}(X \in A) \leq e^{-t^2}, \quad (2.3)$$

and that

$$\mathbb{P}(X \notin A) \mathbb{P}(X \in A) \leq e^{-2(\mathbb{E}[d_\alpha(X, A)])^2}. \quad (2.4)$$

Inequality (2.1) shows that the constant $\frac{1}{2}$ in (1.1) can be improved to 2 for small t . Indeed, when $t \in (0, \mathbb{E}[d_\alpha(X, A)])$, it holds $2(\mathbb{E}[d_\alpha(X, A)])^2 \geq 2t^2$. Thus the bound (2.1) implies that for $t \in (0, \mathbb{E}[d_\alpha(X, A)])$,

$$\mathbb{P}(d_\alpha(X, A) \geq t) \mathbb{P}(X \in A) \leq e^{-2t^2}, \quad (2.5)$$

which concludes our claim.

When $t \rightarrow \infty$, it holds that $(t^2 + (t - 2\mathbb{E}[d_\alpha(X, A)])^2)/(2t^2) \rightarrow 1$. Thus the bound (2.1) also behaves as $\exp\{-2t^2\}$ for large t 's.

Inequality (2.3) shows that the constant $\frac{1}{2}$ in McDiarmid's inequality (1.1) can be improved to 1. The virtue of the bound (2.3) is that it does not depend on $\mathbb{E}[d_\alpha(X, A)]$.

It is well-known that $\mathbb{P}(X \notin A) \mathbb{P}(X \in A) \leq \frac{1}{4}(\mathbb{P}(X \notin A) + \mathbb{P}(X \in A))^2 = \frac{1}{4}$. While, inequality (2.4) implies a stronger result: if $\mathbb{E}[d_\alpha(X, A)] \rightarrow \infty$, then it holds

$$\mathbb{P}(X \notin A) \mathbb{P}(X \in A) \rightarrow 0.$$

This new feature does not imply by (1.1).

Next, we apply Theorem 2.1 to the study of concentration around a median. We have the following improvement on (1.3) and (1.4).

Theorem 2.2. Let X be an \mathbf{E}^n -valued vector of independent random variables. Assume that $\|\alpha\| = 1$, and that the function f satisfies

$$|f(x) - f(x')| \leq d_\alpha(x, x').$$

Let m be a median of $f(X)$. Then for any $t > 0$,

$$\begin{aligned} \mathbb{P}(f(X) - m \geq t) &\leq 2e^{-h(t)} \\ &\leq 2e^{-t^2} \end{aligned} \quad (2.6)$$

and

$$\begin{aligned} \mathbb{P}(m - f(X) \geq t) &\leq 2e^{-h(t)} \\ &\leq 2e^{-t^2}, \end{aligned} \quad (2.7)$$

where $h(t)$ is defined by (2.2) with $A = \{Y : f(Y) \leq m\}$.

Inequalities (2.6) and (2.7) together implies the following non-uniform bound between the medians and the mean.

Theorem 2.3. Assume the condition of Theorem 2.2. Let m be a median of $f(X)$, and let μ be the mean of $f(X)$. Then

$$|\mu - m| \leq (2\rho + \sqrt{\pi/2})e^{-2\rho^2} < 1.5503, \quad (2.8)$$

where $\rho = \mathbb{E}[d_\alpha(X, A)]$ with $A = \{Y : f(Y) \leq m\}$.

Inequality (2.8) shows that the distance between μ and m is decaying exponentially to 0 as $\rho \rightarrow \infty$.

3 Extensions to nonnegative functionals

Let $x \in \mathbf{E}^n$. Denote by $x^{(i)} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \in \mathbf{E}^{n-1}$ which is obtained by dropping the i -th component of x . For each $i \leq n$, denote f_i a measurable function from \mathbf{E}^{n-1} to \mathbf{R} . We have the following extension of McDiarmid's inequality (1.7) to a new class of functions. Such class of functions including the self-bounding functions and the (a, b) -self-bounding introduced by Boucheron, Lugosi and Massart [2] and McDiarmid and Reed [5], respectively.

Theorem 3.1. Let X be an \mathbf{E}^n -valued vector of (not necessarily independent) random variables. Assume that $\|\alpha\| = 1$, and that for some functions f_i , $i = 1, \dots, n$, and all $x \in \mathbf{E}^n$,

$$0 \leq f(x) - f_i(x^{(i)}) \leq \alpha_i. \quad (3.1)$$

Let μ be the mean of $f(X)$. Then for any $t > 0$,

$$\mathbb{P}(f(X) - \mu \geq t) \leq e^{-2t^2} \quad (3.2)$$

and

$$\mathbb{P}(\mu - f(X) \geq t) \leq e^{-2t^2}. \quad (3.3)$$

Recall that a function f is called (a, b) -self-bounding, if

$$0 \leq f(X) - f_i(X^{(i)}) \leq 1 \quad (3.4)$$

and, moreover, for some $a > 0, b \geq 0$ and all $x \in \mathbf{E}^n$,

$$\sum_{i=1}^n \left(f(x) - f_i(x^{(i)}) \right) \leq af(x) + b. \quad (3.5)$$

In particular, a $(1, 0)$ -self-bounding function is known as a self-bounding function. It is easy to see that for an (a, b) -self-bounding function, condition (3.4) implies that $f(X)/\sqrt{n}$ satisfies condition (3.1) with $\alpha_i = \frac{1}{\sqrt{n}}, i = 1, \dots, n$, that is

$$0 \leq \frac{f(X)}{\sqrt{n}} - \frac{f_i(X^{(i)})}{\sqrt{n}} \leq \frac{1}{\sqrt{n}}. \quad (3.6)$$

Thus condition (3.4), the inequalities (3.2) and (3.3) together implies that for any $t > 0$,

$$\mathbb{P}(f(X) - \mu \geq t) \leq e^{-2t^2/n} \quad (3.7)$$

and

$$\mathbb{P}(\mu - f(X) \geq t) \leq e^{-2t^2/n}. \quad (3.8)$$

Notice that the last two inequalities do not assume condition (3.5).

Assume that X is an \mathbf{E}^n -valued vector of *independent* random variables. For any (a, b) -self-bounding function f , McDiarmid and Reed [5] proved that for any $t > 0$,

$$\mathbb{P}\left(f(X) - \mu \geq t\right) \leq \exp\left\{-\frac{t^2}{2(a\mu + b + at)}\right\} \quad (3.9)$$

and

$$\mathbb{P}\left(\mu - f(X) \geq t\right) \leq \exp\left\{-\frac{t^2}{2(a\mu + b + t/3)}\right\} \quad (3.10)$$

See also Boucheron, Lugosi and Massart [2] for the self-bounding functions. Theorem 3.1 does not assume condition (3.5), and holds for *dependent* random variables. Hence, our inequalities (3.2) and (3.3) can be regarded as the generalizations of the inequalities of Boucheron, Lugosi and Massart [2] and McDiarmid and Reed [5].

When X is an \mathbf{E}^n -valued vector of independent random variables, Theorem 3.1 implies the following concentration inequalities around a median.

Theorem 3.2. *Let X be an \mathbf{E}^n -valued vector of independent random variables. Assume that $\|\alpha\| = 1$, and that for all $i = 1, \dots, n$ and all $x \in \mathbf{E}^n$,*

$$0 \leq f(x) - f_i(x^{(i)}) \leq \alpha_i. \quad (3.11)$$

Let m be a median of $f(X)$, and let $\mu = \mathbb{E}[f(X)]$. Then (2.6), (2.7) and (2.8) hold.

4 Proofs of the Theorems

In this section, we devote to the proofs of our theorems.

Proof of Theorem 2.1. Denote by

$$f(X) = d_\alpha(X, A) \quad \text{and} \quad \mu = \mathbb{E}[f(X)].$$

Then it is easy to see that

$$|f(x) - f(x')| \leq d_\alpha(x, x').$$

By McDiarmid's inequality (1.6), it follows that for any $\lambda \geq 0$,

$$\mathbb{E}[e^{\lambda(f(X)-\mu)}] \leq \exp\left\{\frac{\lambda^2}{8}\right\}. \quad (4.1)$$

By McDiarmid's inequality (1.7), we get for any $t > 0$,

$$\mathbb{P}(\pm(f(X) - \mu) \geq t) \leq e^{-2t^2}. \quad (4.2)$$

Since $f(X) = 0$ if and only if $X \in A$, we have

$$\mathbb{P}(A) \leq \mathbb{P}(f(X) - \mu \leq -\mu) \leq e^{-2\mu^2}. \quad (4.3)$$

Thus the last inequality and (4.1) implies that for any $\lambda \geq 0$,

$$\mathbb{P}(A)\mathbb{E}[e^{\lambda f(X)}] \leq \exp\left\{\lambda\mu + \frac{\lambda^2}{8} - 2\mu^2\right\}. \quad (4.4)$$

Using Markov's inequality, we get for any $\lambda \geq 0$,

$$\begin{aligned} \mathbb{P}(A)\mathbb{P}(f(X) \geq t) &\leq \mathbb{P}(A)e^{-\lambda t}\mathbb{E}[e^{\lambda f(X)}] \\ &\leq \exp\left\{-\lambda t + \lambda\mu + \frac{\lambda^2}{8} - 2\mu^2\right\}. \end{aligned} \quad (4.5)$$

When $t \geq \mu$, the right hand side of the last inequality attains its minimum at $\lambda = 4(t - \mu)$. Hence, taking $\lambda = 4(t - \mu)$ in inequality (4.5), we obtain the desired inequality for $t \geq \mu$. When $t \in (0, \mu)$, taking $\lambda = 0$ in inequality (4.5), we obtain the desired inequality for $t \in (0, \mu)$. \square

Proof of Theorem 2.2. Set $A = \{Y \in \mathbf{E}^n : f(Y) \leq m\}$. Then for each $Y \in A$,

$$f(X) \leq f(Y) + d_\alpha(X, Y).$$

Minimising over all $Y \in A$, we have

$$f(X) \leq m + d_\alpha(X, A).$$

Hence, it is easy to see that

$$\{f(X) - m \geq t\} \subseteq \{d_\alpha(X, A) \geq t\}.$$

By Theorem 2.1, for any $t \geq 0$,

$$\begin{aligned} \mathbb{P}(f(X) \leq m)\mathbb{P}(f(X) - m \geq t) &\leq \mathbb{P}(A)\mathbb{P}(d_\alpha(X, A) \geq t) \\ &\leq e^{-h(t)}. \end{aligned}$$

Since $\mathbb{P}(f(X) \leq m) \geq \frac{1}{2}$, we obtain the desired inequality (2.6). Notice that $-m$ is a media for $-f(X)$ and $\mathbb{P}((-f(X)) - (-m) \geq t) = \mathbb{P}(m - f(X) \geq -t)$. Thus

$$\mathbb{P}(-f(X) \leq -m)\mathbb{P}(m - f(X) \geq -t) \leq e^{-h(t)}.$$

By the fact $\mathbb{P}(f(X) \geq m) \geq \frac{1}{2}$ again, we obtain the desired inequality (2.7). \square

Proof of Theorem 2.3. It is easy to see that

$$\mu - m = \mathbb{E}[f(X) - m] \leq \mathbb{E}[(f(X) - m)^+] = \int_0^\infty \mathbb{P}(f(X) - m \geq t)dt.$$

By (2.6), it follows that

$$\begin{aligned} \int_0^\infty \mathbb{P}(f(X) - m \geq t) dt &\leq 2 \int_0^\rho e^{-2\rho^2} dt + 2 \int_\rho^\infty e^{-t^2 - (t-2\rho)^2} dt \\ &= 2\rho e^{-2\rho^2} + 2e^{-2\rho^2} \int_0^\infty e^{-2t^2} dt \\ &= (2\rho + \sqrt{\pi/2})e^{-2\rho^2}. \end{aligned}$$

Thus

$$\mu - m \leq (2\rho + \sqrt{\pi/2})e^{-2\rho^2}.$$

Notice that $-m$ is a media for $-f(X)$, and that $-\mu$ is the means of $-f(X)$. Thus, by (2.7),

$$m - \mu = (-\mu) - (-m) \leq (2\rho + \sqrt{\pi/2})e^{-2\rho^2}.$$

In conclusion, it holds

$$|\mu - m| \leq (2\rho + \sqrt{\pi/2})e^{-2\rho^2} < 1.5503. \quad (4.6)$$

This completes the proof of Theorem 2.3. \square

Proof of Theorem 3.1. Denote by $X = (X_1, X_2, \dots, X_n)$, then $(X_i)_{i=1, \dots, n}$ is a finite sequence of random variables. Let $(\mathcal{F}_i)_{i=1, \dots, n}$ be the natural filtration of the random variables $(X_i)_{i=1, \dots, n}$, i.e. $\mathcal{F}_i = \sigma\{X_j, 1 \leq j \leq i\}$. Let $f(X) - \mu = \sum_{i=1}^n M_i$ be Doob's martingale decomposition of $f(X)$, where

$$M_i = \mathbf{E}[f(X)|\mathcal{F}_i] - \mathbf{E}[f(X)|\mathcal{F}_{i-1}].$$

By condition (3.1), it is easy to see that

$$\begin{aligned} M_i &\leq \mathbf{E}[\alpha_i + f_i(X^{(i)})|\mathcal{F}_i] - \mathbf{E}[f(X)|\mathcal{F}_{i-1}] \\ &= \alpha_i + \mathbf{E}[f_i(X^{(i)}) - f(X)|\mathcal{F}_{i-1}]. \end{aligned}$$

Similarly, we have

$$\begin{aligned} M_i &\geq \mathbf{E}[f_i(X^{(i)})|\mathcal{F}_i] - \mathbf{E}[f(X)|\mathcal{F}_{i-1}] \\ &= \mathbf{E}[f_i(X^{(i)}) - f(X)|\mathcal{F}_{i-1}]. \end{aligned}$$

Notice that $\mathbf{E}[f_i(X^{(i)}) - f(X)|\mathcal{F}_{i-1}]$ is \mathcal{F}_{i-1} -measurable. By Azuma's inequality [1] (see also Corollary 2.7 of [3]), it follows that for any $\lambda \geq 0$,

$$\mathbb{E}[e^{\lambda(f(X) - \mu)}] \leq \exp \left\{ \frac{\lambda^2}{8} \sum_{i=1}^n \alpha_i^2 \right\} = \exp \left\{ \frac{\lambda^2}{8} \right\}.$$

Hence,

$$\mathbb{P}(f(X) - \mu \geq t) \leq \inf_{\lambda \geq 0} \exp \left\{ -\lambda t + \frac{\lambda^2}{8} \right\} = e^{-2t^2}, \quad (4.7)$$

which gives the desired inequality (3.2). Notice that $\mu - f(X) = \sum_{i=1}^n (-M_i)$ and

$$-\alpha_i - \mathbf{E}[f_i(X^{(i)}) - f(X)|\mathcal{F}_{i-1}] \leq -M_i \leq -\mathbf{E}[f_i(X^{(i)}) - f(X)|\mathcal{F}_{i-1}].$$

By an argument similar to that of (4.7), we obtained the desired inequality (3.3). \square

Proof of Theorem 3.2. Denote by $x_{(i)} = (x_1, \dots, x_{i-1}, x'_i, x_{i+1}, \dots, x_n) \in \mathbf{E}^n$ which is obtained by changing the i -th component of x to any given $x'_i \in \mathbf{E}$ such that $x'_i \neq x_i$. Condition (3.1) implies that

$$f(x) \leq f_i(x^{(i)}) + \alpha_i \leq f(x_{(i)}) + \alpha_i.$$

By recursion method, the last inequality implies that for any $x, x' \in \mathbf{E}$,

$$f(x) - f(x') \leq d_\alpha(x, x'). \quad (4.8)$$

Notice that $d_\alpha(x, x') = d_\alpha(x', x)$. By exchanging the places of x and x' , we find that (4.8) is equivalent to that for any $x, x' \in \mathbf{E}$,

$$|f(x) - f(x')| \leq d_\alpha(x, x'). \quad (4.9)$$

Thus Theorem 3.2 follows by Theorems 2.2 and 2.3. \square

References

- [1] Azuma, K., 1967. Weighted sum of certain independent random variables. *Tohoku Math. J.* **19**, No. 3, 357–367.
- [2] Boucheron, S., Lugosi, G., Massart, P., 2000. A sharp concentration inequality with applications. *Random Structures Algorithms* **16**(3): 277–292.
- [3] Fan, X., Grama, I. and Liu, Q., 2015. Exponential inequalities for martingales with applications. *Electron. J. Probab.* **20**(1): 1–22.
- [4] McDiarmid, C., 1998. Concentration. *Probabilistic methods for algorithmic discrete mathematics*. Springer Berlin Heidelberg. 195–248.
- [5] McDiarmid, C., Reed, B., 2006. Concentration for self-bounding functions and an inequality of Talagrand. *Random Structures Algorithms* **29**(4): 549–557.
- [6] Rio, E., 2013. On McDiarmid's concentration inequality. *Electron. Commun. Probab* **18**(44): 1–11.